After the fiesta is over

Mehotcheva, Teodora Hristova

3
The study

As has been discussed in Chapter 1 above, research on language attrition so far has mainly focused on L1 attrition and very little has been done in the area of FL/L2 attrition. The few investigations dedicated to this less popular branch of the field have been targeted at either children (Cohen 1989, Olshtain, 1986), young teenagers (Murtagh 2003, Taura, 2008) or school acquired languages (Weltens 1986, Grendel 1993). It is the aim of the present study to explore FL attrition in adult speakers of a language that has not only been formally acquired but has also been used in real life and to explore the role that factors such as initial proficiency, attitude and motivation, duration of the SA and use and contact with the language might play in the process of retention/attrition. This chapter presents the design of the study. The first section introduces the participants and their sociolinguistic characteristics. This is followed by the materials used: the construction and scoring process are explained and reliability measures, where appropriate are outlined. The design and data collection process are described next. The section closes by outlining the research questions and hypothesis that the study aims at answering.

3.1 Participants

The sample consists of Dutch and German university students who had been on a SA program to a Spanish speaking country. Participating in a SA program meant that the participants had more or less similar experiences with the language and were exposed to similar conditions while in the country, i.e. they followed lectures at the university, had to find accommodation and communicate with their housemates in Spanish, etc. It also meant that the participants had a similar experience with the language before the SA, as university students are usually required to study the language of the destination country before leaving. Although language experiences can never be absolutely the same this ensured a certain level of homogeneity in terms of input, courses taken and materials used to learn the language.

3.1.1 Recruitment and data collection

Recruitment was targeted at university students, present or past, at Dutch and German universities with Dutch or German L1. They had to have participated in a SA program in a Spanish speaking country in the form of an exchange program, Erasmus or similar, or an internship. They could either be still on the SA and were then interviewed at location (interviewing at location was limited to Barcelona due to practical reasons) or they could have already gone back to their country of origin in which case they were interviewed there.

Possible participants were contacted through the international offices and the student advisers at their home universities in the Netherlands and in Germany. Fifteen different universities were contacted but the participants interviewed came only from five institutions: University of Groningen (RuG), Hanze Polytechnic Groningen (Hanze), Radboud University Nijmegen (RU), Technical University Berlin (TUB) and Free University Berlin (FUB). An email message presenting the study, prepared by the researcher, was sent to the international offices of these universities and those willing to collaborate with the project forwarded the message to all students, present or past, involved in a SA with a Spanish speaking country. Then it was up to the interested students to get back to the researcher. Unfortunately, this process was quite cumbersome and the response rate very low. Still, the initial group of participants who expressed the desire to participate in the study helped to recruit more people by word of mouth as they were asked, or offered themselves, to present the project to fellow students or friends of German or Dutch origin who had also spent some time in Spain or another Spanish speaking country. This resulted in a total number of 60 interviewees.

Data collection started in spring 2008 and continued until summer 2009 in three main data collection periods. In the first period, April – June 08, twenty people were interviewed. Five more were interviewed in the second period, November 08. The rest of the participants (35) were interviewed in the third and final period of data collection, April-June 09. These 60 participants were interviewed personally by the researcher at five different locations: Groningen, Nijmegen, Amsterdam, Berlin and

Barcelona. As can be seen in Table 3.1 thirty participants were recruited from Dutch universities, seventeen from Germany universities and thirteen people were recruited by word of mouth and help from friends.

Of the 60 people originally interviewed, nine had to be dropped from the final sample whose data were used in the present analysis. One participant was from Aruba and his mother tongue was Papiamento, a Creole language heavily influenced by Spanish. Three other participants had to be dropped because their L1 was not Dutch or German, but Lithuanian (1) and Bulgarian (2). Three participants had spent twice as much time on a SA as the rest of the participants (24 months). The last two participants did not participate in a SA but in an *au pair* program. Thus, the final sample consisted of 51 participants

Table 3.1 Recruitment of participants by institution, dropped participants and final sample

Origin	*N*
Hanze & RuG	6 & 15
RU	9
FUB & TUB	13 & 4
Other	13
Total recruited	60
Dropped out	9
Final sample	51

All participants in the study were either given a small present, i.e. a box of chocolates, or their name was entered in a lottery with several cash prizes in reimbursement for their participation in the study.

3.1.2 Sociolinguistic characteristics

This section presents the sociolinguistic characteristics such as age, gender and L1 distribution for the whole sample. The duration of the SA and attrition period as well as level of education are reported here. The linguistic background of the participants, i.e. the number of foreign languages that they have been in contact with and/or have studied and their exposure to Spanish at different educational levels is discussed last. In the next chapter, which deals with the CS analysis of the data, the participants are divided into a baseline group and three additional groups depending on length of attrition. Detailed descriptive sociolinguistic

statistics, language experience and background information for each group can be found there.

Duration of SA and LoA

The participants in the study were almost equally distributed between the two sexes, with slightly more men (26) than women (25) (see Appendix A). The mean age was 23.45 (SD 1.8) and it varied between 21 and 29 yeas of age. There were 26 Dutch L1 speakers (51%) and 25 German L1 speakers (49%).

Table 3.2 Duration SA and Attrition

Duration SA	*N (51)*	*LoA*	*N (51)*
		= 0	12 (23.5%)
< 5 months	1 (2%)	1-3 months	9 (17.7%)
5-6 months	35 (68.6%)	3-6 months	2 (3.9%)
7-9 months	8 (15.7%)	7-10 months	17 (33.3%)
10-12 months	7 (13.7%)	11-96 months	11(21.6%)

The mean SA duration (Appendix A) was 6.58 months (SD 2.29) with a minimum duration of 4 months and a maximum 12. The LoA varied from 0 to 96 months, 9.9 months on average (SD 16.3). Table 3.2 shows that most of the participants spent between 5 and 6 months on a SA. For eight people the SA lasted between 7 and 9 months. Seven participants spent between 10 and 12 months. Only one person had been in Spain for less than 5 months (4 to be precise). Regarding LoA, 12 people were at zero months of attrition; 9 had finalized their SA between 1 and 3 months ago; 3 had done so 3 to 6 months ago; 17 had lost contact with Spanish between 7 and 10 months ago and finally, 11 participants had finalized their SA between 11 and 96 months ago.

Level of education

SA usually takes place during the second or third academic year. This automatically excluded first year university students from the sample and meant that the earliest moment participants could be interviewed was towards the end of their second year. Figure 3.1 shows that the majority of the participants were in their fourth year - 11 Dutch and 14 Germans amounting to 49%. Ten people (19.6%), eight Dutch and two Germans

were doing a Master's degree; eight (15.7%) people, four Dutch and four Germans, were in their third year. There were 4 graduates and 4 second year students, (7,84% each) divided between one Dutch and three Germans, both.

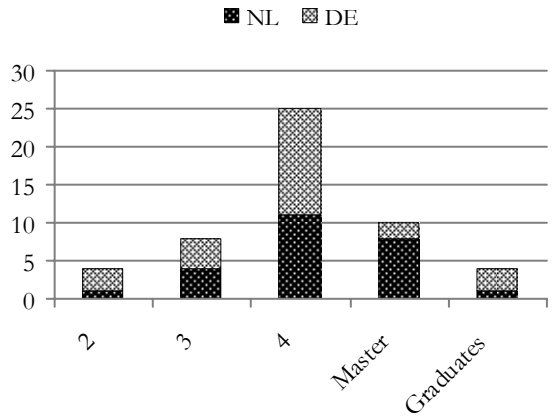

Figure 3.1 Year at university

Linguistic background

The participants in the study had studied a total of seventeen different foreign languages. Figure 3.2 presents the number of languages studied by the participants, depending on their L1. Twenty-two of the participants studied 4 foreign languages, with a slight advantage for the L1 Dutch participants (22 to 14). Two, six and seven languages were studied by two people each, with one L1 Dutch and one German L1 speaker for the first two and only Dutch L1 speakers for the last one. Nine German L1 speakers had studied three foreign languages and one Dutch L1 participant five.

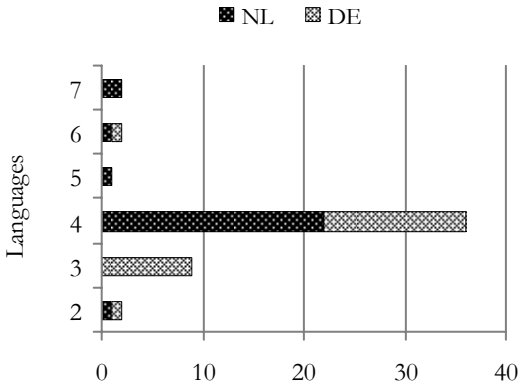

Fig 3.2 Number of foreign languages studied

Spanish was L5 for the majority of the participants – 69.3% including their L1; L4 for 17.3%; L3 for 3.84% and L6 for 9.61%. The majority of the participants had studied Spanish at University for approximately two years (84.2%); 13.5% started studying Spanish as early as secondary school and 69.2% had Spanish classes while on the SA. These consisted of the free language classes which were provided for participants in the Erasmus program and varied from 30, 40 to 60h classes. The estimate of ~40h is based on the median number of classes reported by the students, who were not always a hundred percent sure about the number of classes that they had attended. Only one person studied Spanish after the SA.

3.2 Materials

In an attempt to avoid the problem of differences in methodology which was discussed in the previous chapter, the present study followed the methodology and the language attrition test battery developed by Schmid (2005). The battery, suggesting various tests and measures, from oral interviews and film retelling to picture naming and grammaticality judgment tasks, has already been applied in a number of investigations (Altenberg & Vago, 2004; Keijzer, 2007; Ribbert & Kuiken, 2010; Schmid & Dusseldorp (2010); Tsimpli, 2007).

Not all of the questionnaires and tests suggested by Schmid were considered appropriate for the present study and the ones used also had

to be adapted for the needs of the investigation. Five different tasks were employed to collect data (see Figure 3.3).

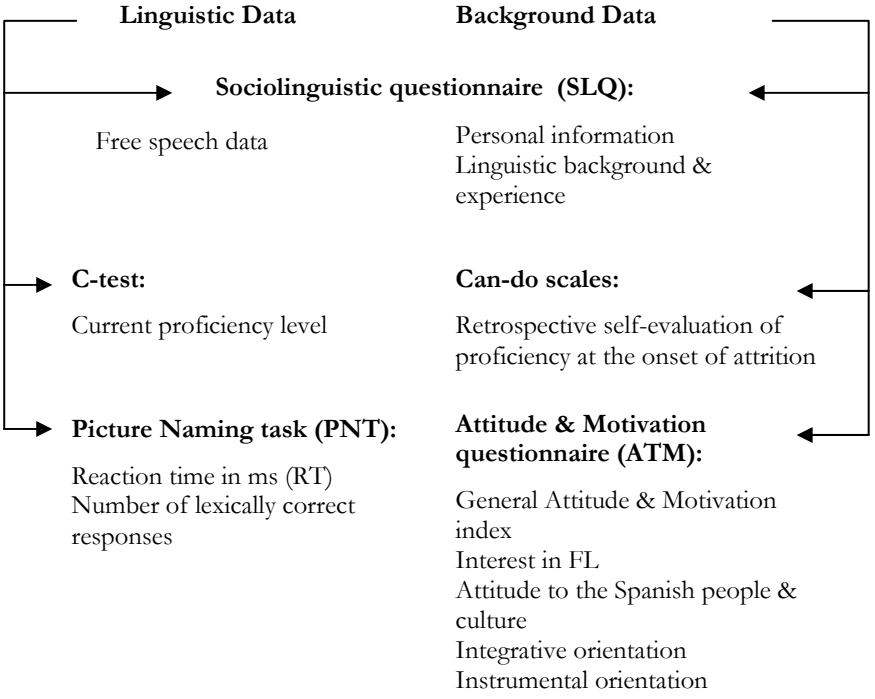

Figure 3. 3 Data collection materials

There were three instruments to gather linguistic data: the sociolinguistic questionnaire (SLQ) which was used as the basis for an interview eliciting oral data in addition to the background sociolinguistic and language contact and use information; a C-test with a focus on nouns and a picture naming task (PNT). Two other questionnaires, attitude and motivation questionnaire (AMQ), based on Gardner (1985), and can-do scales (following the Common European Framework of Reference for Languages - CEFR) were used to gain more insight into the participants' attitudes and motivation to learn foreign languages and Spanish in particular, and their initial level of proficiency, i.e. before they lost active contact with the language.

3.2.1 Sociolinguistic questionnaire

The SLQ (SLQ) was designed to be used in a semi-structured interview. A semi-structured interview allows for the introduction of new questions in the course of the interview as a consequence of what the interviewee says. The advantage of a such an interview, in comparison to a retelling task, is that it allows the interviewer to direct the conversation to a topic that the participant finds more appealing and is familiar with, which gives the participants a chance to perform at their best. In a story or a film retelling task, people are limited to the vocabulary required by the task, which they might not be familiar with or not feel confident using. Also, it was felt that a free conversation was the closest to the way the people involved in the project used Spanish while on a SA, which, is characterised by a large amount of oral practice and input and lack of written practice.

Aim and previous use of oral data in attrition studies

The SLQ was used to provide both sociolinguistic and oral data as it was conducted as a semi-structured interview. The oral data was then used to access lexical diversity, hesitation and pause phenomenon. The measures that were adopted in the present study were based on previous research on attrition (Schmid & Beer Fägersten, 2010). They focused on the frequency and distribution of disfluency and hesitation phenomena (Table 3.3), namely: filled pauses, false starts, corrections, repetitions and reformulations.

Table 3.3 Oral data measures

Frequency disfluency markers:	*Distribution disfluency markers*	*Lexical Diversity*
• Filled Pauses	• Art	
• False Starts	• Adj	• D
• Corrections	• Adv	
• Repetitions	• Noun	
• Reformulations	• Etc.	

Besides evaluating their frequency, disfluency markers were examined for their distribution, i.e. whether they appeared predominantly in front of a particular part-of-speech element or were evenly distributed. If lexical access was compromised, it was expected that this might be manifested

in an increase of hesitation and disfluency phenomena preceding lexical items. Another measure that was calculated on the basis of the oral data was lexical diversity. Lexical diversity measures the range or richness of vocabulary used by a person. A speaker who uses a limited selection of words has low lexical diversity, in contrast to a person who uses lots of synonyms and diverse vocabulary to express herself. Previous research has found these measures to be especially relevant to the study of language attrition. A study by Schmid & Beers Fägersten (2010) has demonstrated that disfluency phenomena can change in the course of L1 attrition and that the position of disfluency markers may signal not only lexical retrieval difficulties but also point to problems with specific grammatical features.

Constructing the test

The SLQ (SLQ) was based on the Language Contact Profile developed by Freed, Dewey, Segalowitz & Halter (2004) and the language background questionnaire developed by Andonova[1] and used at the Central and East European Center for Cognitive Science, Sofia, Bulgaria. The questionnaire contained a total of 19 questions (see Appendix L) and consisted of three sections: Personal information and linguistic background, SA experience and Current linguistic experience.

The first section, Personal information and linguistic background gathered information such as age, place of birth, education, languages used by the parents and within the family and languages that the participants had studied at different points in their lives. The participants also had to self-evaluate their abilities in each of these languages that they had studied and mark their contact with Spanish at different educational levels, i.e. pre-school, primary education, secondary education, etc. The questions were given as open ended questions (see example below) in the course of the talk and the researcher was in charge of noting down the answers.

5. ¿Qué lengua(s) hablas en casa?
Holandés ☐ *Alemán* ☐ *Otra* _____

What language do you speak at home?
Dutch☐ *German*☐ *Other* _____

[1] Personal communication

Section two, SA Experience, gathered information about the participant's experience and living arrangements during the SA as well as the frequency with which the participant used their languages to perform different activities while in Spain. The questions about the living arrangements were brought up during the conversation and the researcher marked the answers. Then the participant was in charge of filling in the table for language use and frequency (an excerpt can be seen bellow).

17. Por favor, utiliza la escala siguiente para marcar la frecuencia y la lengua que utilizabas ***durante tu estancia en España***.

1– muy raramente; 2 – raramente; 3 – a veces; 4 – con frecuencia; 5 – todo el tiempo

Cuando:	español	inglés	holandés/alemán	Otra
hablaba con amigos				
hablaba con mascotas				

Please use the following scale to note the frequency and language that you were using during your SA in Spain:

1 – almost never 2 – rarely 3 – sometimes 4 – frequently 5 – always

When:	*Spanish*	*English*	*Dutch/German*	*Other*
talking to friends				
talking to pets				

Section three, Current linguistic experience, was a repetition of the languages and frequency part of section two but this time the participant was instructed to answer it on the basis of the period *after* coming back from the SA. Again, it was filled in by the participant. This section was not administered to people who were still on a SA.

Administering the questionnaire

The questions from the SLQ were used as a basis for the interview, and although obligatory, they were not exclusive. Thus, depending on the interests and mood of the participant, the interview could go to one direction or another and cover additional topics to the ones included in the questionnaire. It was considered that allowing the participants to talk about things that they liked and were close to, rather than confining them to a certain picture story or a set topic, would help them relax, feel more confident and perform at their best.

Preparing the data for analysis

In order to analyze the data from the interview, all free speech samples were recorded on a digital recorder and then transcribed using the conventions of the Codes for the Human Analysis of Transcript (CHAT). CHAT together with CLAN (Computerized Language Analysis) are two basic components of the CHILDES project developed by MacWhinney (2000) which are used for the analysis of conversational interactions. Although the project was originally developed to analyze child language, hence the name CHILDES – Child Language Data Exchange System, CHAT and CLAN are increasingly used for the analysis or adult speech as well (Pérez-Vidal et al., forthc.; Valls & Mora, 2009; Yilmaz, van der Kooi & Schmid, 2009). Transcribing and storing the data in CHAT format made it possible to use a number of programs from the CLAN package such as FREQ – providing frequencies of specific or all items in the corpus; MOR – performing a syntactic tagging of the transcribed text, and VOCD – calculating the type/token ratio. These programs will be discussed in more detail later on in the section.

To ensure the correct functioning of the CLAN programs, the data first had to be transcribed conforming to the CHILDES standards. An example of a transcribed interview can be found in Appendix N. Besides transcribing the data in the necessary format, certain codes, i.e. for hesitation, repetition, reformulation were introduced. Part of the coding was done simultaneously with the transcription by introducing specific codes for the disfluency phenomena. For a full list of all symbols used in the transcriptions see Appendix M.

Filled pauses, irrespective of their pronunciation, i.e. um, ah, uh, were all coded with the 'ahm@fp' symbol as in the following example:

1) um el febrero de uh dos mil siete um hasta ah junio /
   ahm@fp el febrero de ahm@fp dos mil siete ahm@fp hasta ahm@fp junio

   um February uh two thousand and seven um until ah June.

Repetitions of a part of a word were coded as *false starts* with the '&" symbol in front of the incomplete item as in Example 2 below.

2) pues en Málaga esta &ta también más difícil porque < tienen > [/] tienen un acento muy fuerte

well in Malaga it's &al also more difficult because they < (they) have> (they) have a very strong accent

Repeating linguistic material in the same way without any correction or change was coded with the symbol [/] following the repeated material in angled brackets. This could include single words, Examples 2, 4 & 5, or strings of two or three of words as in Example 3.

3) y mi barrio también, como bueno < no es muy > [/] no es muy, malo pero no esta bien, sabes?

and my neighbourhood, well, <it's not very> [/] it's not very, bad but it's not good, you know?

Corrections, or partial repetition of the preceding linguistic material with a correction, were coded with the symbol [//] following the retracted materials in angle brackets as in Examples 4.

4) sí, pero en Kenia y Tanzania es lo < mujer > [//] < mejor > [/] mejor, sí.

yes, but in Kenya and Tanzania it's the <bast> [//] <best> [/] best, yes.

Finally, *reformulations* of the preceding linguistic material were coded with the symbol [///] following the reformulated material in angle brackets.

5) algo así, < había > [///] pues era una casa < de > [///] < con > [/] con diferentes familias y todo

something like that, <there was> [///] well, it was a house <of> [//] with different families and all

Originally, it was intended to mark empty pauses as well and in the first stages of the coding process this was done manually by the researcher with the intention to later have native speakers independently mark silent pauses and compare the inter-rater agreement. However, it became clear that a lot of the pauses were caused not by linguistic problems and searching for the right word but by trying to remember when, where, why, etc. something was done. This was confirmed by the native speaker who started marking pauses independently of the researcher and claimed not to be able to clearly distinguish between linguistic-based pauses and other pauses.

Defining a pause in the speech of a FL speaker seems to be a very challenging task. How do we know when a pause is related to language processing problems or simply to trying to retrieve a memory? Especially so in the present study where the interview focused on a past experience. In pausological research a distinction is made between short and long pauses (see, for example, Kormos & Dénes, 2004). But how long is a long pause in the speech of a FL speaker? While 'norms' as to the speed, rate and pausing patterns of native speakers and fluent L2 speakers can be easily obtained, how do we establish this for the FL speaker? Since exploring this problem is beyond the scope of the present study, it limits itself only to the analysis of filled pauses.

The files were then morphologically tagged by means of the MOR routine developed by Brian MacWhinney and Monica Sanz Torrent. After doing the MOR analysis, the output was disambiguated by means of the POST routine. The POST program uses a database file which contains information about syntactic order for the respective language and it is part of the grammar package. All further analyses such as counting frequencies, exploring the position of hesitation and disfluency markers or calculating lexical diversity were carried out on the files produced by the POST command. A complete transcription after the MOR and POST programs had been run can be found in Appendix O.

3.2.2 C-test

A C-test is a modified version of the Cloze test designed on the principle of reduced redundancy but avoiding some of the pitfalls that Cloze tests pose. In the C-test, instead of deleting whole words (as in the Cloze test) only the second half of every second word is deleted. This on the one hand allows for a much higher deletion rate (allowing for shorter texts to be used) and on the other hand (usually) leaves only one possibility for a correct answer making the task of both the tested and the scorer much easier. C-tests have been shown to be "the most economical and reliable procedure" (Klein-Braley, 1997:47) among the reduced redundancy tests and also to be "a reliable and valid procedure representative of the reduced redundancy principle" (Babaii & Ansay, 2001, p. 209).

Aim and previous uses of the C-test in attrition studies

The C-test was used as a means of assessing general language proficiency and it was designed with a special focus on nouns so that noun production could be compared with the PNT. C-tests have been used in a number of studies on L1 attrition (Keijzer, 2007; Köpke et al., 2007; Yagmur 1997) as well as in studies on L2 attrition (Murtagh, 2003) to establish the participants' level of proficiency.

Constructing the C-test

When creating a C-test, creators and investigators Raatz & Klein-Braley (1998, point 4.2.1) recommend using texts that:

- are written texts complete in themselves
- are appropriate in difficulty and content for the target group
- have no specialised vocabulary or content
- are not literary texts or contain verbal humour

Three different texts that met the above mentioned conditions, were taken from the materials for the official exam of Spanish as a foreign language (DELE) from the Cervantes Institute webpage[2]. They were put into C-test format with a deletion rate of every third word rather than every second word to avoid a disproportionally high percentage of deleted articles and pronouns. This also allowed to put the stress on the production of nouns which were also investigated with a psycholinguistic task so that a comparison could be made in the performance under two different conditions. The first and the last sentences of each text were left intact. One-letter words such as "y" (*and*) were ignored as well as personal names and names of places. In words with uneven number of letters, one more letter was deleted than left standing. Each deleted letter was represented by a line.

The three texts were then pre-tested with 10 native speakers of Spanish and 5 fluent foreign speakers of Spanish currently living in Spain. As a result, two of the texts had to be discarded because they were two

[2] http://diplomas.cervantes.es/candidatos/recursos.jsp

difficult[3]. The scores for the text that was left (see Appendix P) were 97% for the natives and 91% for the non-native speakers. Table 3.4 shows the distribution of parts-of-speech in the final text, which had 58 gaps.

Table 3.4 Percentage original and deleted words

Category	*% deleted words*	*% original words*
Nouns	25.86%	21.26%
Verbs	24.14%	18.55%
Adjectives	8.62%	5.88%
Adverbs	10.34%	6.78%
Prepositions	6.89%	10.85%
Conjunctions	6.89%	10.40%
Articles	10.34%	12.66%
Pronouns	5.17%	8.59%
Contractions	1.72%	1.35%

Administration

C-tests have been previously employed in a study on the attrition of Irish as a second language (Murtagh, 2000) and in a study on L1 attrition in Dutch immigrants in Anglophone countries (Keijzer, 2007; Köpke et al., 2007; Yagmur 1997). In both cases, however, the participants had a time limit of 5 mins per text, whereas in the present investigation no time limit was imposed. That was done on purpose as it was felt that it might be too frustrating for the FL attriters. Also, the participant was instructed to try and reconstruct as many words as possible but not to worry if something was left unfinished as it was almost impossible to reconstruct all words. If the participant had not finished after 10 mins, the researcher told them that that was enough and asked them to move on to the next questionnaire.

[3] Native speakers should get at least 95%correct restorations (Raatz & Klein-Braley, 1998)

Scoring the C-test

Although Raatz & Klein-Braley (1981) advocate that, in order to avoid problems of subjectivity caused by the scorer's judgment of what is acceptable and what is not, only exact scoring should be used, it was felt that a group of attriters who were moving away from the linguistic norm were quite different from a group of SL learners who were aspiring to get closer to it. A misspelled attriter's answer can still be quite an achievement whereas in the case of a SL learner, it can be a mistake. In order to limit the possibility of personal interpretation of what constitutes a correct answer and still be able to explore the type of errors made by the participants, a 7 point scale similar to the one suggested by Schmid (2005) was used: *0* – if the gap was left empty; *1* – incorrect stem and incorrect word class; *2* – incorrect stem but correct word class; *3* – correct stem, incorrect word class; *4* – agreement error, be it number agreement, tense agreement, etc; *5* – all the previous are ok but still something is wrong; *6* – correct choice with a spelling mistake; and *7* – correct.

Table 3.5 Type and frequency of spelling mistakes

correct word	*spelling error*	*frequency*
además	ademas	13
canción	cancion	4
dolía	dolia	7
más	mas	14
niños	ninos	7
pequeño	pequeno	7
podía	podia	4
anécdota	anécdote, anéctoda	14, 1
llevaba	llevada, llemaba	1, 1
porque	porqua	1
privado	privada, privato	1, 2
quires	quiras, quienes	1, 1
tuvó	tubo	1

The score that was calculated for each participant (0-58) was based on codes 6 and 7 and the rest of the scale was used only for qualitative analysis of errors. Table 3.5 lists the occurrences of misspelled words. As can be seen, the spelling mistakes, more often than not, consisted of omissions of diacritics and very rarely there were misspelled words.

Reliability

Reliability measures for the C-test were obtained with Cronbach's α. Cronbach's α measures the internal consistency of the items within a questionnaire or whether the different items of a test measure the same thing. It is based on the principle of splitting the data in half and is essentially a correlation for all possible split combinations. Eight items (3, 9, 13, 19, 21, 30, 43 and 53) were excluded from the reliability analysis as they had zero variability. Table 3.6 shows that the overall reliability of the test was high, α=.89 Although it was not as high as the reliability reported in Murtagh (2003:80), i.e. α=.94, it was well within the 'unofficially' accepted limit of .7 - .8 (Field, 2005).

Table 3.6 Reliability analysis of the C-test: Cronbach's α

N items	*N cases*	*min; max score obtained*	*Mean*	*SD*	*Alpha*
50	51	29;57	46.10	7.484	.889

Item difficulty

Table 3.7 shows descriptive item difficulty statistics for the gaps by parts-of speech category and for the total. Item difficulty was calculated on the basis of percent correct reconstructions per item, therefore a high percentage signaled low difficulty and vice versa. The average success rate was 79.4%, considerably closer to the higher end of the scale indicating that the test was not too difficult for the attriters. Verbs obtained the lowest percent correct restorations, while prepositions the highest.

It is not surprising that CONJ and PREP obtained a high success rate. They are two-letter words requiring only one letter for restoration (see Appendix Q for the number of letters to be restored in each gap). This applies to ART as well, as there was only one restoration that required three letters and 5 one-letter reconstructions, making the task relatively easy. It should be noted that, in terms of difficulty, nouns come in second, after verbs, but with considerably higher success rate. Appendix Q shows individual item difficulty ratings for each word. Item difficulty rate varied from 0 to 100%.

Table 3.7 Percent correct responses for part-of-speech category and total.

	N	*Min*	*Max*	*Mean*	*SD*
N	15	0	100	79.6	26.9
V	14	29.4	94.1	66.7	17.7
ADJ	5	76.5	98.0	88.2	8.7
ADV	6	37.3	100	77.5	26.0
CONJ	4	80.4	100	94.1	9.1
PREP	4	88.2	100	94.1	5.7
ART	6	76.5	100	83.7	8.3
PRO	3	60.8	94.1	77.8	16.6
CONTR	1	88.2	88.2	88.2	-
Total	58	0	100	79.5	20.4

3.2.3 Picture naming

A picture naming task (PNT) with 75 stimuli was used to obtain reaction time measures and proportion of lexically correct responses for all participants.

Aim and a PNT as a measure in attrition

The aim of the PNT was to explore the processes of lexical access in FL attrition. The analysis of the PNT generated two measures: *1*) reaction time (RT) analysis, i.e. the time it took the participant to name the picture stimulus and *2*) proportion of correct responses (PCR), i.e. the percentage of correctly named pictures irrespective of reaction times. Although being a very popular tool in research on bilingualism, timed picture naming (PN) has been used only in a small number of language attrition projects like Ammerlaan's (1996), Hulsen's (2000) and Soesman's (1997) on L1 attrition. Although Schmid & Köpke (2009) note that not all of these studies used reliable reaction time measurement equipment, they maintain that PN is a valid measure for exploring problems with lexical retrieval in attriting populations. Recently, a PNT (alongside other measures) has also been adopted in a large scale investigation on L1 attrition in Moroccan and Turkish immigrants in the Netherlands (Van der Kooi, Yilmaz & Schmid, 2009; Yilmaz, van der Kooi & Schmid, 2008). To the knowledge of the researcher, however,

the present project would be the first one to employ a PNT in the study of FL attrition.

Constructing the PNT

The PNT consisted of 75 pictures, controlled for three levels of frequency, shown on a computer screen. Cognate words with Dutch and German were excluded. In addition there were 10 practice pictures. The criteria for choosing the stimuli, the apparatus used and the procedures followed are discussed in the following sections.

Stimuli

The stimuli for the experiment consisted of 75 black-and-white drawings taken from two different sources: 61 from Sanfeliu & Fernandez's (1996) set of "245 Snodgrass-Vanderwart pictures standardized for Spanish" and 14 from the *On-line Resource for Psycholinguistic Studies* (Szekely et al., 2004), developed and maintained by the University of California at San Diego. This was necessary since the on-line resource database, although allowing to browse for pictures using different parameters such as semantic category, percent name agreement and length of syllables, was standardized for Mexican Spanish. A comparison between the two sets showed that there were some differences as to the Name agreement and Image agreement ratings, maybe due to cultural bias and differences. Although these differences were not large, it was decided to follow the norms standardized for Spanish as far as possible and to use the Mexican norms only to fill in for any necessary items. A complete list of the stimuli with their origin can be found in Appendix R and pictures of the experimental stimuli are shown in Appendix T.

As can be seen from Table 3.8, the stimuli used represented different semantic categories such as people, animals, body parts, objects, foods, etc. Culturally biased drawings such as a football helmet, a raccoon or a skunk that are more common within North American culture; antiquated objects like *a spinning wheel*, *a thimble*, *a top* or technical vocabulary items like *a chisel*, *pliers*, *screw*, *screwdriver*, were excluded from the selection of the stimuli.

Table 3.8 Distribution of semantic categories across stimuli (excluding trial stimuli) based on the On-line Resource for Psycholinguistic Studies (Szekely et al., 2004).

Semantic categories	*N*
People	5
Animals	18
Body parts	7
Vehicles	3
Foods	7
Things to wear	4
Small artefacts	21
Large artefacts	6
Objects or nature phenomena	4
Total	*75*

- *Name agreement*

One of the conditions that the stimuli for the experiment had to meet was to have a minimum 80% name agreement. Name agreement (NA) is the degree to which participants agree on the name of a drawing. Drawings that generate several different names have lower name agreement than pictures generating only one. NA has been demonstrated to be a good predictor of naming speed (Barry, Morrison & Ellis, 1997; Vitkovitch & Tyrell, 1995). Drawings that elicit only one name, i.e. *dog,* are named faster and more accurately than drawings that can elicit more than one, i.e. *gun,* which can also be named *pistol* or *revolver.*

Table 3.9 Name agreement (NA) and image agreement (IA)

	Stimuli	*Min*	*Max*	*Mean*	*SD*
NA	75	80	100	95.71	5.127
IA	75	3	4.47	3.86	.422

NA is calculated as the percentage of people that produce the target name and the statistic *H* that was suggested by Snodgrass & Vanderwart (1980). Although at first sight it seems that these two factors measure the same thing, the statistic *H* provides more information as to the distribution of names across participants. For example, as Snodgrass & Vanderwart (1980, p.184) explain 'if two concepts both are given their dominant name by 60% of the participants, but one is given a single other name and the second is given four other names, both concepts will have equal percentage agreement scores, but the first will have a lower *H* value'. The

choice of the stimuli for the study was based on the % NA and the statistic *H* was only monitored in case of a low % (<85%) NA. The mean NA for the stimuli was 95.71 (StD= 5.127) (see Table 3.8).

- *Image agreement*

Image agreement (IA) refers to the degree to which the mental image that a participant forms when presented with a name corresponds to the actual picture. Barry et al. (1997) demonstrated that pictures with higher ratings of IA had shorter naming latencies than pictures with lower ratings. They suggested that IA has its influence at the level of object recognition, so that the closer a picture is to one's mental image of an object, the faster the naming for that item will be. IA ratings were taken from Sanfelui & Fernandez (1996). They were based on students' rating of the image they formed to a sound stimulus over 7secs (3 after hearing the word and 4 after seeing the picture) and rated on a 5 point Likert scale, 1 being low agreement and 5 high agreement. Only stimuli with minimum level 3 of IA were considered for the study. The mean IA for the stimuli was 3.86 (StD=.422) (see Table 3.8).

- *Frequency*

As was discussed in Section 2.3 above, word frequency influences naming latencies. In order to assess such frequency effects in naming latencies and in the number of correct responses, three levels of frequency were distinguished. The ratings were based on the International Picture Naming Project Database (Szekely et al., 2004). There were 25 pictures with high frequency – HF (word frequency 5.400, SD 1.1013), 25 pictures with medium frequency – MF (word frequency 3.473, SD 0.4342) and 25 pictures with low frequency – LF (word frequency 2.029, SD .6557). This frequency was also matched to the frequency given by Alameda & Cuetos (1995) in the *Diccionario de frecuencias de las unidades lingüísticas del castellano*. The frequency values between the three sets of pictures were significant at p<.000 level for all three sets.

- *Other characteristics*

In the 75 stimuli there were four complex words (arco iris [rainbow], palomitas [popcorn], paraguas [unbrella] and tijeras [scissors]. The ratio between masculine/feminine words was 45.3% to 54.&% and the ratio of animate/unanimate, 33.3% to 66.7%.

Trial stimuli

The trial stimuli came from the same sources as the experimental stimuli in a 8/2 ratio: Spanish set and Mexican set (see Appendix R for the stimuli origin). To avoid priming, none of the stimuli included in the trial appeared later in the experiment script. The trial consisted of 10 drawings that were representative of the semantic categories present in the experiment and that always appeared in the same order: *star [estrella], banana [platano], church [iglesia], coat [abrigo], moon [luna], pencil [boligrafo], bridge [puente], boot [bota], boy [niño]* and *bear [oso]*.

Randomized lists

In order to diminish the tiredness effect, i.e. the fact that participants tend to become slower towards the end of an experiment because of fatigue, four different randomized versions were created. The order of appearance of the words was controlled for semantic category and initial sound of the word. That is, two consecutive words could not belong to the same category or begin with the same sound.

Apparatus

E-prime version 1.1.4 was used to create and run the script. All experiments were carried out on an ASUS X51R series portable computer. The screen was a 15,4" WXGA with a 1280x800 pixel resolution and a refresh rate 59.905 Hz. The black and white digitalized drawings (300x300 pixels) were shown in the middle of the screen. The participants had a hand-held microphone that was connected to a Serial Response Box which measured the reaction time (RT) with a voice key.

Administration

Before starting with the experiment the participants were instructed to name the picture they were going to see on the screen as quickly as possible, using only one word without an article and avoiding extra-linguistic sounds such as "hm", "ahh", etc. They were told to remain silent if they could not name the word (See Appendix S for the full instruction). In order to familiarize the participants with the task and the stimuli they could expect, a short trial version was run before the actual experiment. It could be repeated several times if the participant did not feel confident to start with the experiment. It also served to see whether the microphone was triggered correctly by the participant's voice or its sensitivity level had to be adjusted.

Once the trial session was over and the participant confirmed that she was ready to continue with the actual experiment, she could indicate this by means of pressing the space bar. On each trial, first there was a fixation cross "+" that appeared centered on the screen for 1000 ms. This was followed by the stimulus which disappeared as soon as the voice key was triggered (or for a maximum of 10.000ms[4]) and "*" appeared on the screen for 1000ms signaling voice-detection. The period between the offset of one trial and the onset of the next one was a random value of between 1200ms and 1500ms to prevent participants from falling into a pattern.

During the experiment the researcher used a score sheet (one for each of the 4 randomized versions) to follow the participant's progress. A six-point scale was used to mark problems with the RT (*1-target word with a valid RT; 2 – target word with a false start, hesitation, self correction; 3 – target words early RT; 4 – target word with late RT; 5 – target word and no RT; 6 – No Response, no RT).* In order not to make the participants nervous they were informed beforehand that the researcher might write down numbers and that these concerned only the way the microphone worked. Any names that did not match the target were noted down later on from the digital recording of the experiment, again in order not to make the participants anxious that they were not doing well.

[4] Pre-testing of the experiment started with 3000ms response time, that was increased to 4000ms and finally left to 10000ms as attriters tended to name the word when the stimulus was already gone and the next one was on.

Scoring

Two different scores were calculated on the basis of the PNT: mean reaction times (RT) in ms and percent correct responses (PCR).

The RT analysis was based on a 6-point scale (shown in Table 3.10 below) which was also used by the researcher during testing to code reaction time related events for each item on the list.

Table 3.10 Reaction time codes

Code	*Description of the code*
1	the word produced matched the target word and RT was detected correctly
2	the word produced matched the target word but there was a false start, hesitation, self correction
3	the word produced matched the target word but RT was registered too early
4	the word produced matched the target word but RT was registered too late
5	the word produced matched the target word but the mic failed to register a response
6	no response was given and no RT was registered

There were 3 codes (codes 3, 4 and 5, which signaled malfunctioning of the microphone and 3 codes (codes 1, 2 and 6) which were related to the response given by the participant. The most frequent one was Code 1 – for a correct response with a valid reaction time measure. It accounted for 59.1% of all RT codes. The second most frequent code was Code 6 (no response given) with 30.2%. Failure of the microphone to trigger when there was a valid response (Code 5) added up to 5%. Early and late triggering of the microphone, code 3 and 4 respectively, contributed with 2% each and finally, words where the participant stuttered, hesitated or there was a false start amounted to 1.6%.

For the % correct responses another, lexical, code was introduced, which was independent of the RT code and which evaluated the lexical correctness of the response given, i.e. matching the target, synonym, hypernym, hyponym, etc. Table 3.11 lists the lexical codes used. These were based on the 7-point scale used by Bates al. (2003), i.e. codes 1 to 7, plus five additional codes (8-12), which were added to distinguish between the different types of errors. Items which did not get a response were coded with 0. Adding a lexical code allowed to analyze responses

that were excluded from the RT analysis because of problems with the microphone, i.e. those marked with codes 2, 3, 4 and 5, in the percent correct responses analysis.

Table 3.11 Lexical codes, based on Bates et al. (2003)

Code	*Code description*	*target*	*response*
1	the response matches the target	gato (*cat*)	gato (*cat*)
2	morphological variation with phonological overlapping, the truth value is preserved	casa (*house*)	casita (*little house*)
3	synonyms	sombrero (*hat*)	gorro (*hat*)
4	hypernym	tiburon (*shark*)	pez (*fish*)
5	hyponym	araña (*spider*)	tarantula (*tarantula*)
6	word in a foreign language, the truth value is preserved	fresa *(strawberry –es*)	fragola (*strawberry-it*)
7	wrong answer	rana (*frog*)	gusano (*worm*)
8	morphological variation of gender, word incorrect	molin*o* (*windmill*)	molin*a*
9	morphological variation of number, word incorrect	paragua*s* (*umbrella*)	paragua
10	semantic association	avion (*airplane*)	vuelo (*flight*)
11	pseudo word based on the target	zapato (*shoe*)	zapatines
12	pronunciation error	calcetin /kalθetin/ (*sock*)	calceton/kalθeton/ (*sock*)

The responses were coded by the researcher and independently by another rater who was a native speaker and a teacher of Spanish, using the lexical codes from Table 3.11. An inter-rater reliability analysis using Cohen's Kappa statistic (Cohen, 1960) was carried out to determine consistency between raters. Cohen's Kappa is a statistical measure of inter rater agreement, which in addition to calculating the percent agreement, calculates the amount of agreement that can result by pure chance. Although Kappa has been criticized for being too conservative and sometimes underestimating agreement, it is recommended over simple percent agreement, which can be misleading and too liberal (Lombard, 2004). Kappa ranges between 0 and 1 with larger values indicating better agreement. Generally a Kappa >.70 is considered to show substantial agreement. The inter rater agreement for T1 was found to be K = .71 and for T2 K = .86 showing satisfactory inter rater agreement for both data collection times.

3.2.4 Can-do scales

Can-do scales have been used in a number of L1 attrition studies (Hulsen, 2000; Van der Kooi, Yilmaz & Schmid, 2008) to measure the self perceived proficiency of participants and have been demonstrated to be a good measure of second language proficiency.

In the present study they were used as a means of investigating the participants' language proficiency before the onset of attrition., i.e. as a retrospective pretest. The length of time elapsed since the participants' return from the SA (onset of attrition) varies across the sample (ranging from 0 to 8 years ago) and there was no objective way of establishing the participants' proficiency (with the exception of the baseline group) at that time. Instead, the participants from the AG were asked to self-rate their ability to perform different activities in Spanish before the end of their stay in Spain by means of can-do scales.

The use of retrospective pretests, or then-tests, was first suggested by Howard (1980) as a way to control for response-shift bias occurring in conventional pretest-posttest designs. The validity of the retrospective pretests has been investigated in a number of studies (Hoogstraten 1982, 1985; Howard, 1980) and retrospective pretests have been shown to be an accurate and valid measure.

Constructing the can-do scales

The can-do scales questionnaire was based on the ALTE Can Do project of the Common European Framework of Reference (see Appendix U). It consisted of 52 statements that referred to an array of different actions, performed in different contexts and varying in difficulty. The statements covered the reading, writing, listening and speaking skills from level A1 to level C2.

Scoring the can-do scales

The participants had to self-rate their ability to perform each of the actions mentioned using a five-point Likert scale. The response options varied from '*I cannot do this at all*' to '*I can do this without any difficulty at all*' and were represented by numbers from 1 to 5. The maximum score was

260 and the minimum 52. The mean over all items was calculated and the closer it was to 5, the higher evaluation the participant had marked.

Reliability

Reliability of the can-do scales, measured again with Cronbach's α, showed that the questionnaire had a good total reliability of α = .96 (Table3.12). The highest reliability coefficient was measured in the Speaking Performance section. Writing performance obtained α=.92. Listening Comprehension reliability was a little bit lower, α=.85. The lowest reliability was found in the Reading Comprehension.

Table 3.12 Mean scores and reliability coefficients for the can-do-Scales.

	N	*Min; Max*	*M*	*SD*	*a*
Listening Comprehension	10	10; 50	38.8	4.92	.85
Reading Comprehension	9	9; 45	31.8	5.15	.83
Speaking Performance	21	21; 105	70.25	12.02	.94
Writing Performance	12	12; 60	39.15	9.29	.92
Total	52	52; 255	180.89	27.64	.96

No official reliability measures were found for the ALTE framework although according to the information on ALTE's webpage[5], validation work has been in progress for quite some time now.

3.2.5 Attitude and motivation questionnaire

Attitude towards Spanish speaking people and culture, and motivation to learn Spanish were studied with the help of the AMQ. The questionnaire was created using the Attitude Motivation Test Battery (AMTB) developed by Gardner (1985). Not all of the nineteen measures of the original AMTB were included in the AMQ since scales like the Parental Encouragement and the Class Anxiety were not relevant to the present study.

[5] http://www.alte.org/downloads/index.php?doctypeid=10

Constructing the AMQ

The AMQ consisted of 30 questions taken from four of the original AMTB sections: interest in foreign languages (11 questions), attitude towards the Spanish people – 9 questions; integrative orientation – 5 questions; instrumental orientation – 5 questions (See Appendix V for the questionnaire). These were translated into Spanish by the researcher and checked by a native speaker and teacher of Spanish. The participants were instructed to mark to what extent they agreed or disagreed with each statement referring to the Spanish language, culture or people, or language learning. The answers were marked by means of a seven point Likert scale where the response options varied from "Strongly Disagree " to "Strongly Agree" and were represented by numbers from 1 to 7.

Scoring the AMQ

A high score on the AMQ, maximum score 140 (70 interest in foreign languages and 70 attitude towards Spanish people and culture) indicated positive attitude and high motivation to learn the language. A high score on the instrumental orientation section, maximum score 35, showed that the participant had instrumental reasons (finding a job, better salary) for learning Spanish and a high score on the integrative orientation section showed that the participant was motivated by integrative reasons (interest in the culture and language studied). Before scoring the AMQ, the scores for the four inverted items (4, 9, 21 and 25) were reversed in SPSS using the formula: [(max value + 1) - the actual score]. Thus a person with a positive attitude who scored 1 on a negatively worded item would still score 7 and vice versa.

Reliability

Reliability was again tested with Cronbach's α. Table 3.13 shows that the lowest reliability coefficient in the AMQ was measured in the section 'Attitude towards Spanish people', α=.74. The other three section obtained similar coefficients. Integrative orientation had α value of .77; instrumental orientation α=.78 and the highest α value was found in interest in foreign languages, α=.79. These fall well within the ranges obtained by Gardner (1985) although not at the maximum end. It has to be noted that *1*) Garnder's study was conducted with Junior high school

and High school students, aged 12 – 18; and *2*) the number of cases in the present study (n=51) is relatively small and it might be influencing negatively the reliability analysis.

Table 3.13 Mean scores and reliability coefficients (Cronbach α) for the A&M sections. IFL – interest in foreign languages; ASP – attitude towards Spanish people; IO – integrative orientation and InstO – instrumental orientation

AMQ	*Present Study*					*Alpha ranges Gardner (1985)*
	N items	*Min; Max*	*Mean*	*SD*	*Alpha*	
IFL	11	11; 77	66.80	.994	.79	.72 - .90
ASP	9	9; 63	45.44	.879	.74	.67 - .94
IO	5	5; 35	30.69	.496	.77	.62 - .88
InstO	5	5; 35	22.81	.692	.78	.13 - .77

3.3 Design of the study

It was argued earlier that the two major challenges for any study on FL language attrition were establishing a baseline against which to compare attriting individuals and using data collection materials that provide for a multi-dimensional look into the problem. It was seen in the previous section that the materials which were used were quite diverse and allowed for a multidimensional investigation. In this chapter, the design used to overcome the first problem is presented.

Using longitudinal (LG) designs to overcome the 'baseline' problem faces several obstacles: mainly lack of time and the inability to go back and/or forth in time and interview the participants just before the onset of attrition and then a few years later without having to wait for a couple of years. Instead, Weltens (1987:27) suggested using informants as close to the profile of the attriting group as possible and for whom the attrition process has not yet started, i.e. a baseline group. Following his advice, the present study used a baseline group (n=14) of Erasmus participants who were interviewed shortly before the end of their SA (while still in Spain or within a month after going back to their country of origin) and an attriting group (n=37). In addition, as data collection continued for an year, it was decided to re-interview the participants who were interviewed in the first stages of the project (n=20) a year later in order to provide LG data for this subsample.

Figure 3.4 shows the three data collection times: T0, T1 and T2, the people interviewed and the tasks used at each data collection point. The first data collection time, Time 0 (T0), was not a 'real' data collection time. It did not chronologically precede T1 and it consisted only of the can-do scales: retrospective for the attriting group and at-the-time-of-interview for the baseline group. These questionnaires were actually administered together with all other materials at T1 but their data was used as T0 to compare the initial levels of proficiency.

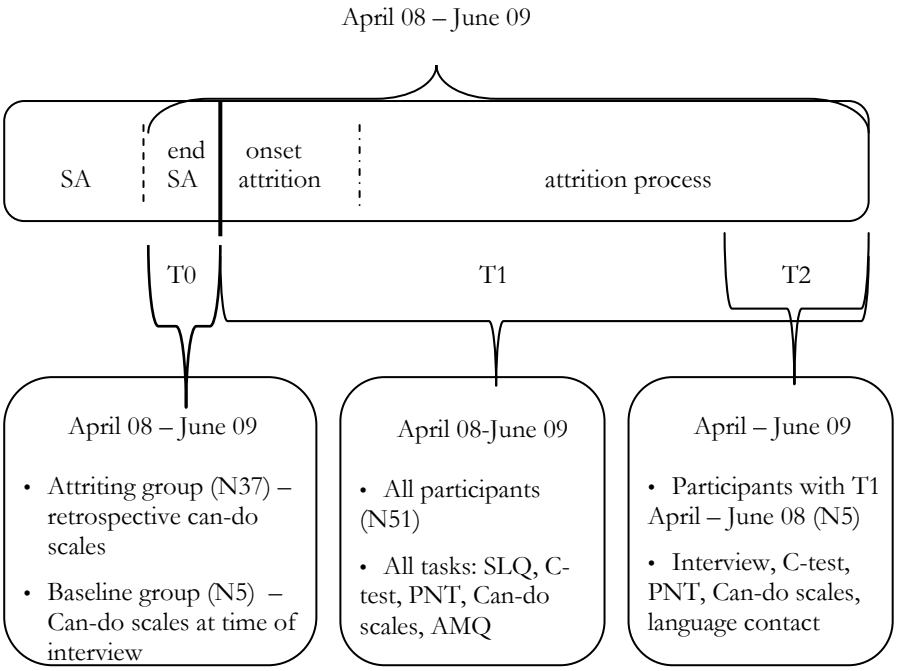

Figure 3.4 Data collection times

Time 1 (T1) data collection, in which participated all interviewees - attriters and baseline group, began in spring 2008 and continued through early summer 2009. All tasks were administered at this data collection time to a total of 51 people.

At Time 2 (T2), only a subsample of the participants who were interviewed in spring 2008 were re-interviewed approximately a year

after their first interview. Originally, at T2 it was intended to re-interview all participants from the early stages of T1 data collection. Unfortunately, a year later the majority of these people had already finished their university education and were either very busy with their new job or where living outside the country. Thus, from the 20 people that were interviewed in spring 2008, only 5 could be retrieved for a follow up interview a year later. T2 for the 5 participants who were re-interviewed continued from April until June 2009.

3.4 Procedure

All participants were interviewed individually by the researcher at the university premises (RuG, FUB and UPF), in an informal setting. Spanish was the language of communication and whenever the participants found difficulty they were encouraged to look for the Spanish words rather than change to English. A typical interview had the following sequence:

1. Consent form
2. SLQ (as an interview)
3. AMQ
4. Can-do scales
5. C-test
6. PNT

At a typical interview at T1 the participant was first asked to sign a consent form (Appendix W). This was the only document written in English to ensure its understanding by everybody irrespective of their level of proficiency in Spanish. The participant was then offered a cup of tea or coffee and the meeting began with an informal talk with the researcher in which she gradually introduced the questions from the SLQ.

The interview was then followed by the rest of the materials which the participant completed on their own, although the researcher was always available to answer questions and help with any doubts. The psycholinguistic task came last. The purpose of having the interview, the questionnaires and the tests, that were all done in Spanish, before the

psycholinguistic task was to "warm up" the participant and activate their Spanish before the most demanding task – the picture naming.

Usually, after the PNT the conversation would continue for a while discussing the participant's impressions of the task and how she felt while naming the pictures so that the interview did not end too abruptly. This also provided valuable insight as to how the participants felt during the task and the problems they encountered such as tip of the tongue situations, confusion with other languages, etc. The whole meeting continued for approximately 1,5h to 2h and it was recorded from the moment the participant entered the room until she left. The interview was later transcribed using the conventions of CHAT (see Chapters 4 and 5 for the analysis of the data).

The materials used at T2 included the language use section from the SLQ, can-do scales, C-test and the PNT. The interview was based on information from the first interview discussing what happened in the meantime and the future plans of the participant. This session was shorter, about 1h 15 min to 1h 30 min, and was again recorded and the interview transcribed for the analysis of free speech.

3.5 Hypotheses and expectations

In addition to the research questions brought up in the previous chapter, the study hopes to confirm a number of hypothesis and expectations again based on the theories discussed in the previous chapter but also specific to the tasks and design used in the study. These are presented separately for each task, while the hypotheses regarding the factors affecting attrition come last.

3.5.1 Oral data

1) In the LG data a decrease in lexical diversity within subjects over time will be observed. There will be a decrease across groups in the CS data; the baseline group will obtain the highest result and the group with the longest LoA the lowest.

2) In the LG data, there will be an increase in disfluency phenomena over time. In the CS data, hesitation and disfluency markers will increase across groups in comparison to the baseline group.

3) The increase in disfluency markers will be mostly visible in front of lexical items.

3.5.2 C-test

4) Scores on the C-test will decrease over time for the LG subsample and across groups for the CS data.

3.5.3 Picture naming

5) Individual naming latencies will increase and percent correctly named words will decrease over time for the LG group. In the CS data, reaction times will increase and percent correct responses will decrease across groups; the baseline group will perform best, i.e. attain a higher percentage of correct words and faster naming latencies, whereas the other groups will obtain lower scores and slower naming latencies; groups with longer attrition periods will have lower scores and slower naming latencies.

6) High frequency words will be retained better and retrieved quicker. There will be more correctly named HF words than MF and LF words. HF words will also be named faster than MF and LF words. This frequency effect will be present in both the CS and LG data.

3.5.4 Factors affecting the attrition process

7) High initial proficiency in the language fosters retention, i.e. people with high proficiency at the onset of attrition will retain the language better and will perform better at the test and tasks.

8) Motivation, especially integrative motivation, will have a positive effect on language retention, i.e. people with high motivation will retain the language better.